THE ELECTRONICS RESURGENCE INITIATIVE

TOM RONDEAU

PROGRAM MANAGER
DARPA/MTO

DOMAIN-SPECIFIC SYSTEM ON CHIP (DSSOC) PROGRAM

THE HISTORY OF GNU RADIO INSPIRING THE KEY TECHNICAL AREAS OF DSSOC

Intelligent Scheduling | Domain Representation | MAC | Software

2001 Single core, single thread (Pentium 4)

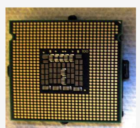
commons.wikimedia.org

+

www.mccdaq.com

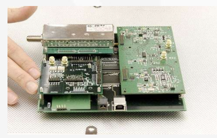
PCI data acquisition card

2004 Growing with technology (Pentium D)



commons.wikimedia.org

+



media.wired.com

Gen 1 Universal Software Radio Peripheral

2008 Multi-threaded scheduling (Cell Broadband Engine)

commons.wikimedia.org

2008 Moving to multi-core processors (i7)

commons.wikimedia.org

www.ettus.com

Gen 2 Universal Software Radio Peripheral

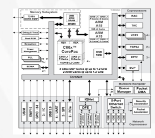
2013 Embedded computing (Arm and Xilinx) Introduction of VOLK

libvolk.org

VOLK

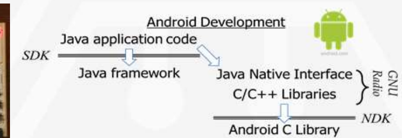
2014 Porting to other SoCs (TI Keystone II)

processors.wiki.ti.com



www.ti.com

2015 Porting to the smart phone (Android devices)



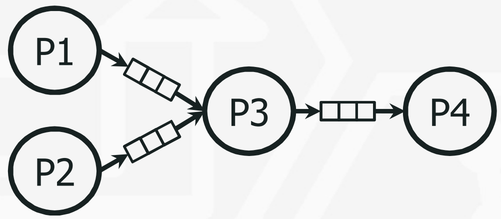
Today

Processors	Computer	I/O
Intel *Arm* *PowerPC*	*Server farms* *Workstations* *Laptops* *Smart Phones* *Embedded Systems*	*USB 2.0* *USB 3.0* *Thunderbolt2* *1 GigE* *10 GigE* *PCIe x4*

MULTI-THREADED SCHEDULING IS A RESOURCE MANAGEMENT ISSUE

Intelligent Scheduling | Domain Representation | MAC | Software

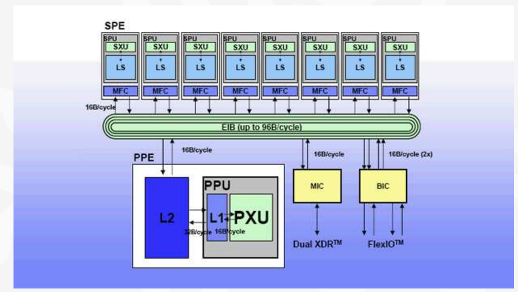
Scheduling processes on parallel resources is dependent on the access structure and programming design



2008 Multi-threaded scheduling (Cell Broadband Engine)

commons.wikimedia.org

- Heterogeneous cores
- Ring buffer to dispatch data to the SPEs
- Local memory

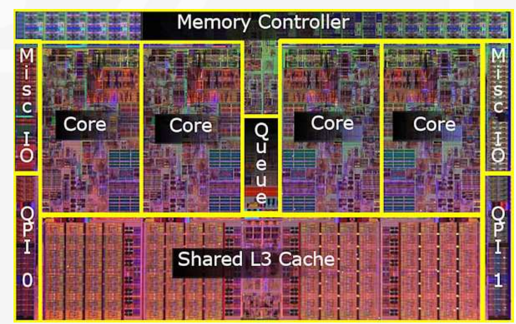


www-03.ibm.com

2008 Moving to multi-core processors (i7)

commons.wikimedia.org

- Homogeneous cores
- Common bus access
- Shared cache and memory access

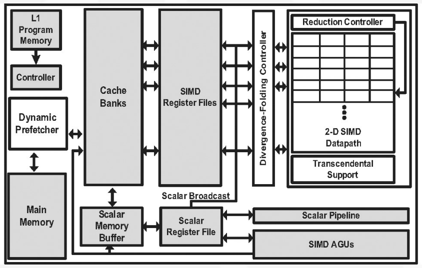


www.legitreviews.com

IDENTIFYING AND MAKING USE OF EXISTING SPECIALIZATION

Single Instruction, Multiple Data (SIMD) accelerators are very close to the core processor; data movement is not the problem

Intel Architecture

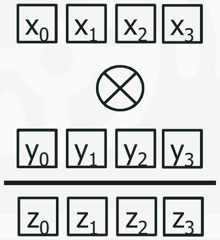


doi.ieeecomputersociety.org

SSE Assembly

_mm_load_ps
_mm_mul_ps
_mm_load_ps
_mm_store_ps

SIMD Concept



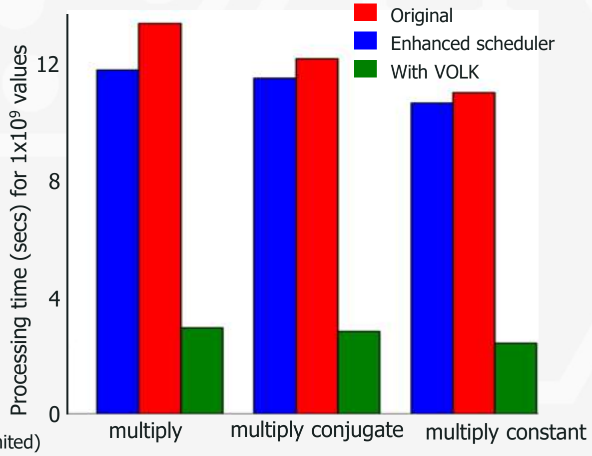
2013 Embedded computing (Arm and Xilinx) Introduction of VOLK

libvolk.org

The Vector-Optimized Library of Kernels (VOLK) was introduced in 2013 to handle SIMD operations

Currently supports:
- SSE, SSE2, SSSE3
- SSE4.1 and SSE4.2
- AVX and AVX2
- NEON (ARM v7 and v8)
- Altivec

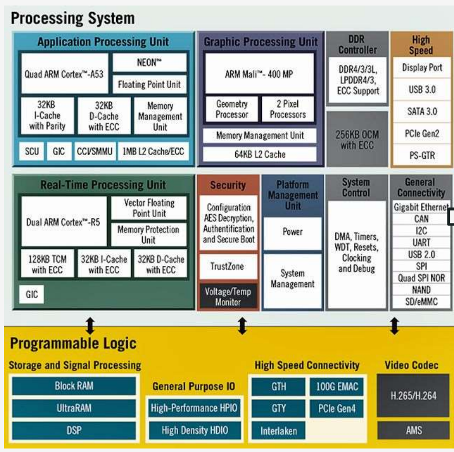
GNU Radio Speed-up with VOLK



Maximizing use of the processor capabilities was critical for GNU Radio to run on embedded systems

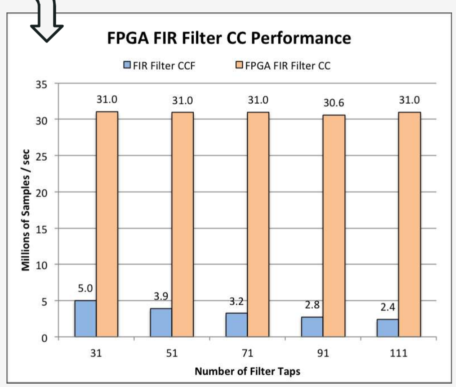
SPECIALIZATION ONLY MATTERS IF ACCELERATORS ARE ACCESSIBLE

2013 Embedded computing (Arm and Xilinx)



xlnx.i.lithium.com

Google Summer of Code (GSoC) project successfully integrated Xilinx Zynq's ARM and Programmable Logic together to more efficiently execute filters

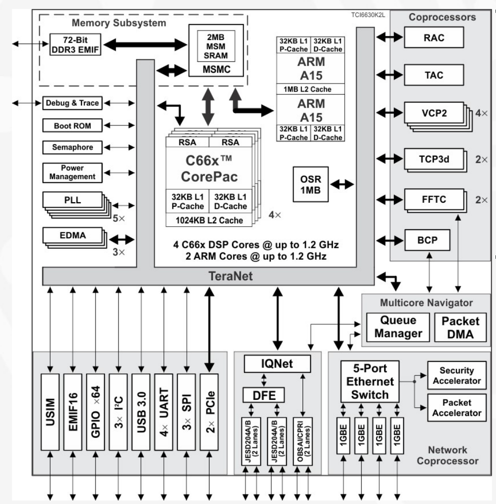


gnuradio.org

Xilinx AXI bus proved usable and efficient

2014 Porting to other SoCs (TI Keystone II)

processors.wiki.ti.com



Accelerators were over-specified to LTE

www.ti.com

- GSoC 2014 looked at using the TI Keystone II SoC
- Failed at managing the CPUs, DSPs, and hardware accelerators across TerraNet
- Data structures and access was unmanageable

DEVELOPERS NEED PROGRAMMING TOOLCHAINS, DEBUGGING, AND PERFORMANCE MONITORING

Intelligent Scheduling | Domain Representation | MAC | **Software**

2015 Porting to the smart phone (Android devices)

Software radio on tables and smart phones is a really exciting form-factor

Android Development

android.com

Java application code

SDK

Java framework

Java Native Interface

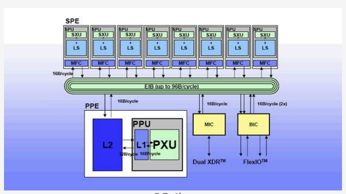
C/C++ Libraries

GNU Radio

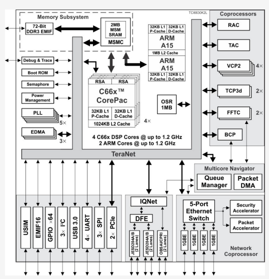
NDK

Android C Library

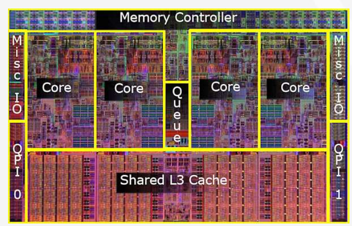
Figuring out how to use it effectively is… less exciting



www-03.ibm.com



www.ti.com



www.legitreviews.com

Hard to program; new compilers; bad debuggers; worse performance analysis tools

Intel improves hardware with the same developer tools and libraries we already knew how to use

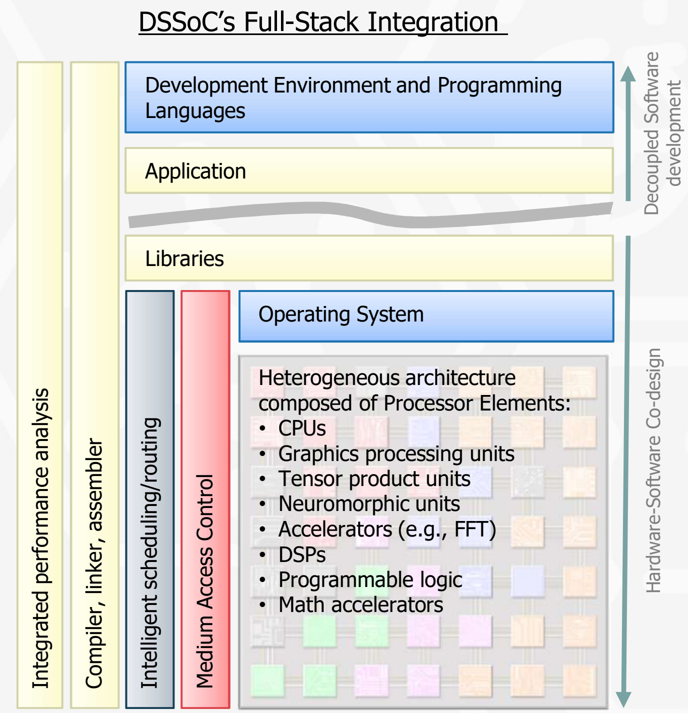
GETTING THE BEST OUT OF SPECIALIZATION WHEN WE NEED PROGRAMMABILITY

Three Optimization Areas

1. Design time
2. Run time
3. Compile time

Addressed via five program areas

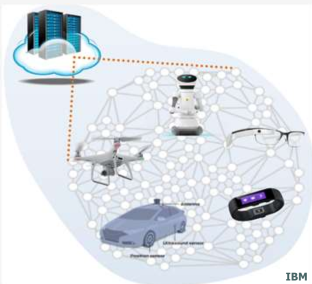
1. Intelligent scheduling
2. Domain representations
3. Software
4. Medium access control (MAC)
5. Hardware integration



Hardware/Software co-design is an enabler for efficient use of processing power

DSSOC SELECTED PARTICIPANTS

IBM T. J. Watson Research Center
Pradip Bose
Columbia University, Harvard University, Univ. of Illinois at Urbana-Champaign



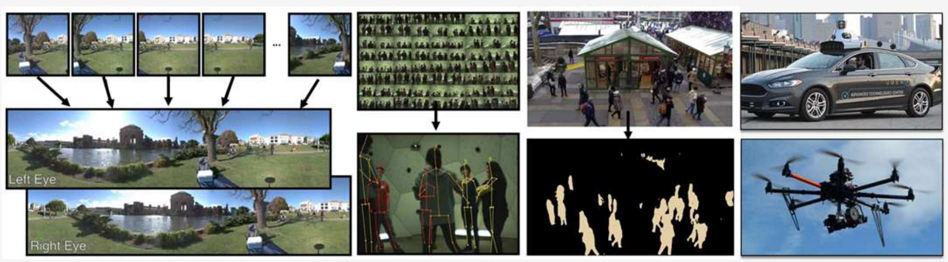
CV+SDR
- Multi-domain application
- Multi-spectral processing
- Communications

Arizona State University
Daniel W. Bliss
Univ. of Michigan, Carnegie Mellon University, General Dynamic Mission Systems, Arm Ltd., EpiSys Science

SDR
- Unmanned aerial
- Small robotic & leave-behind
- Universal soldier systems
- Multifunction systems

Stanford University
Mark Horowitz

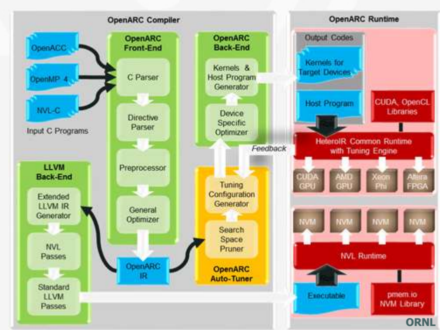
Computer Vision
- Still image and video processing
- Autonomous navigation
- Continuous surveillance
- Augmented reality



Stanford

Oak Ridge National Laboratory
Jeffrey Vetter

SDR
- Communications and signal processing focused
- Up-front processing / data cut down
- Improving understanding of processing systems



Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

ERI ELECTRONICS RESURGENCE INITIATIVE

SUMMIT

2018 | SAN FRANCISCO, CA | **JULY 23-25**

PRADIP BOSE

DISTINGUISHED RESEARCH STAFF MEMBER AND MANAGER OF EFFICIENT & RESILIENT SYSTEMS

IBM T.J. WATSON RESEARCH CENTER

AI OUT TO THE EDGE: CHALLENGES IN SOFTWARE-HARDWARE CO-DESIGN

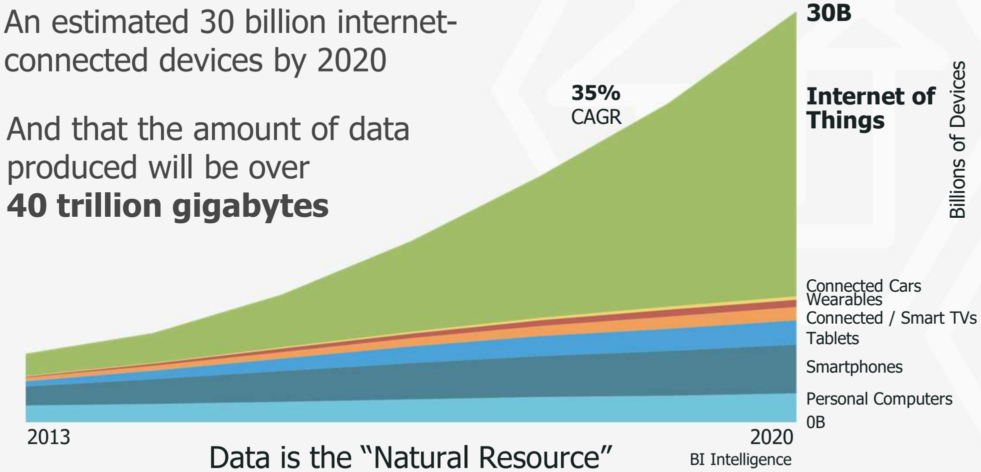
JEFFREY L. BURNS, PH.D.

IBM T. J. WATSON RESEARCH CENTER

AI: WHY NOW? WHAT IS DIFFERENT THIS TIME?

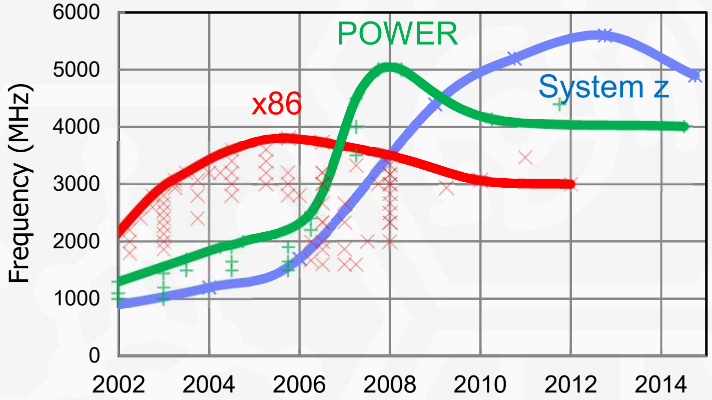
An estimated 30 billion internet-connected devices by 2020

And that the amount of data produced will be over **40 trillion gigabytes**



Data is the "Natural Resource"

Programmable systems cannot keep pace with the volume, complexity and unpredictability of information in the modern world



- Frequency, IPC scaling saturation drove transition to a throughput-centric, multi/many-core model
- Increasing core count will saturate as well, due to power and area limits
- ***AI performance and efficiency demands require specialization***

A NEW ERA OF COMPUTING...

Artificially Intelligent Systems learn and interact naturally with people to amplify what either humans or machines could do on their own

They help us solve problems by penetrating the complexity of Big Data

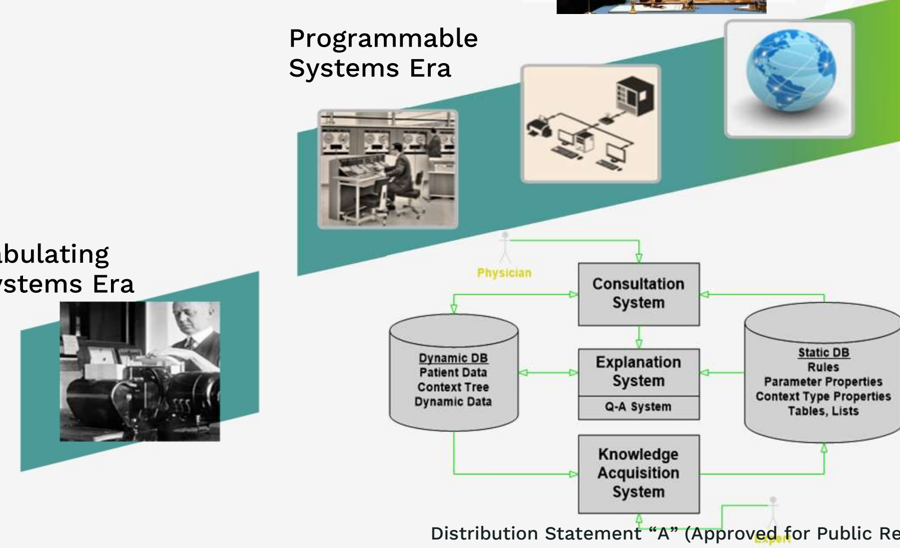
AI Systems Era

Deep Blue

Watson - Jeopardy

Programmable Systems Era

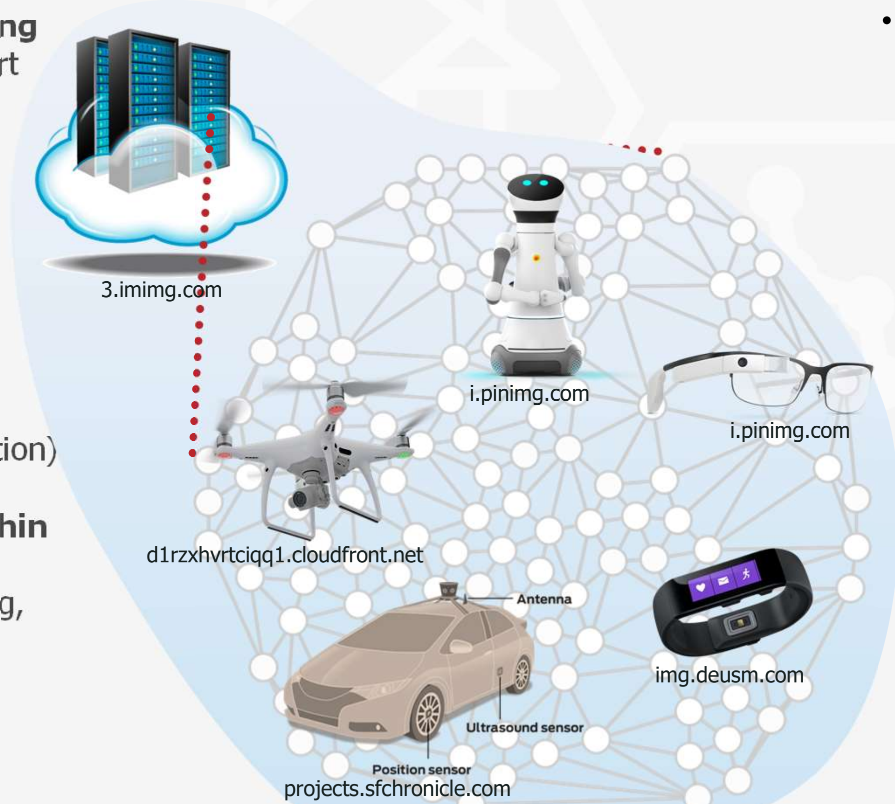
Tabulating Systems Era



Rules-based Expert Systems (e.g. MYCIN)

A NEW SYSTEM ARCHITECTURAL VISION FOR THE COGNITIVE IOT ERA

- **Mobile (swarm) computing**
  - With on-demand support from cloud
- **Unstable wireless bandwidth**
  - Interaction over ad hoc networks
- **Resilient system reconfiguration**
  (on node failure or idle rotation)
- **Adaptive abstraction within devices**
  - Approximation, sampling, filtering
  - Machine learning acceleration
  - Dynamic voltage and frequency control

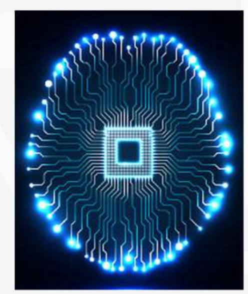


The domain of mobile cognition

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

- **Needs at / near the edge:**
  - On-device inference
  - On-device training
  - Low power / voltage (possibly harvested energy)
  - Harsh environment resilience
  - Security against attacks

Custom cognitive hardware with built-in resilience features



EFFICIENCY GAP: A GRAND CHALLENGE IN COMPUTING

Applications

Cloud PCs Smartphones Wearables

Increased compute and data requirements

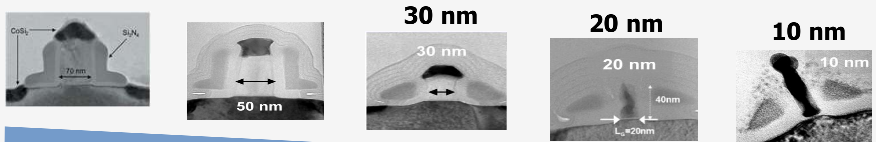
*Need **new sources of efficiency** across the computing stack to bridge the gap!*

(Software-Hardware Co-Design)

- Algorithm and Software
- Compiler and OS
- Architecture
- Logic and Circuits

Efficiency gap

30 nm 20 nm 10 nm



Scaling limits?

Diminishing benefits from technology scaling

Materials & Devices

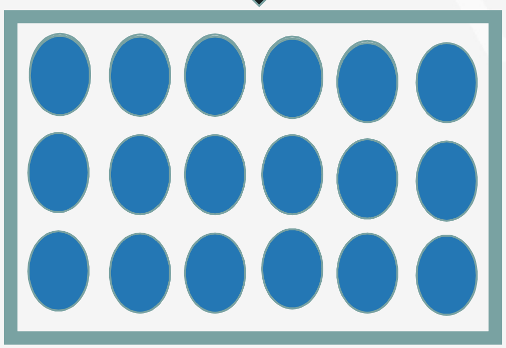
Courtesy: Swagath Venkataramani, IBM

ACHIEVING HIGH PERFORMANCE AND ENERGY EFFICIENCY: SIGNIFICANT TECHNICAL CHALLENGE

Low Voltage Operation

Good for many-core parallelism throughput/watt

- Low-Vmin SRAM
- Aggressive under-volting

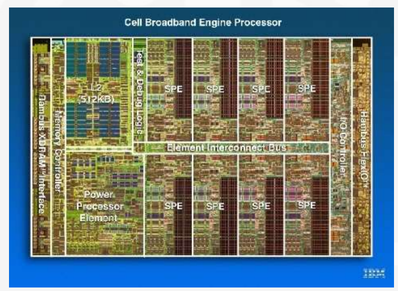


But single thread performance *(hence higher voltage, frequency)* important too!

Wide operating range processors?

Specialization – Accelerators

Improves latency of specialized functions at lower energy cost

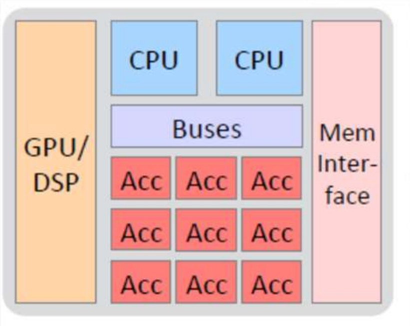


IBM Cell Processor
- First mainstream heterogeneous multi-core
  - *Game machines*
  - *Supercomputers*
- But challenging to program!

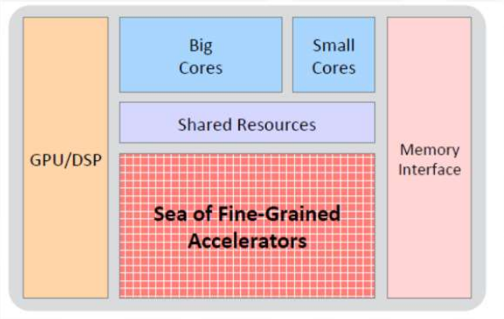
Low-voltage resiliency and approximating computing are promising additional knobs to bridge the gap – *results of research in IBM-led, DARPA-sponsored project under PERFECT program*

EDGE EMBEDDED INTELLIGENT SOC TRENDS

Today's SoC



Tomorrow's SoC



Challenges:

Design integration complexity, validation, efficiency, reliability, programmability, …

DARPA ERI DSSoC Program

(Tom Rondeau, Program Manager)

EPOCHS – IBM'S DSSOC PROJECT

- Incredible growth in R&D investment and public excitement around **connected autonomous cars**

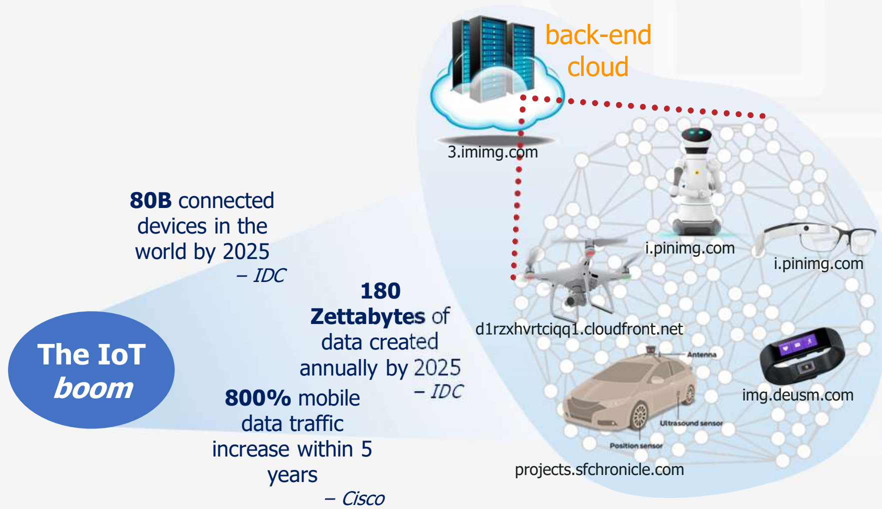
  - **2M+ fully autonomous cars on the road by 2025**
    Source: BI Intelligence
  - **Each car will generate about 40TB of data for every 8 hours of driving**
    Source: Intel

- **EPOCHS:** ***Efficient Programmability Of Cognitive Heterogeneous Systems***
- Rapid development of multi-application systems via a single programmable, heterogeneous SoC
- Focus on **Intelligent Connected Autonomous Vehicles** with **computer vision** and **software radio** applications

EPOCHS: EFFICIENT PROGRAMMABILITY OF COGNITIVE HETEROGENEOUS SYSTEMS

Domain: Cognitive IoT at the Edge

Strategic for DoD and Commercial



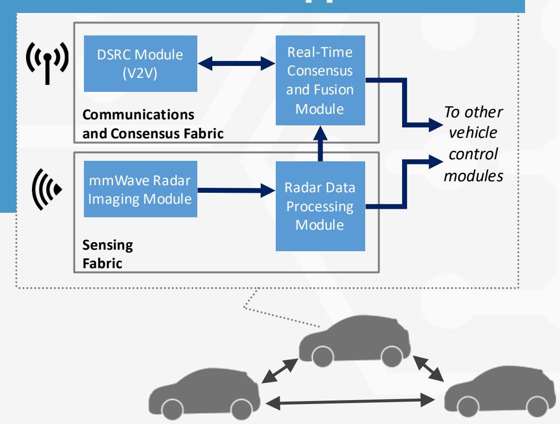
- Motifs / Algorithms:
  - FFT, WAMI, other SDR/CR primitives
  - Image / video / audio analytics
  - Machine Learning / Deep Learning

- PI: Pradip Bose (IBM)
- Co-PIs: Sarita Adve, Vikram Adve, Sasa Misailovic (UIUC)
- Co-PIs: Luca Carloni, Ken Shepard (Columbia)
- Co-PIs: David Brooks, Gu-Yeon Wei (Harvard)

- IBM Research and Harvard expertise & assets:
  - *Low power, resilient ASICs (DARPA PERFECT)*
- Columbia University expertise & assets:
  - *Embedded Scalable Platforms (DARPA PERFECT)*
- UIUC expertise & assets:
  - *LLVM open-source compiler; memory system arch*

ERA: EPOCHS Reference Application

Multi-vehicle cooperative sensor fusion application to drive rapid development

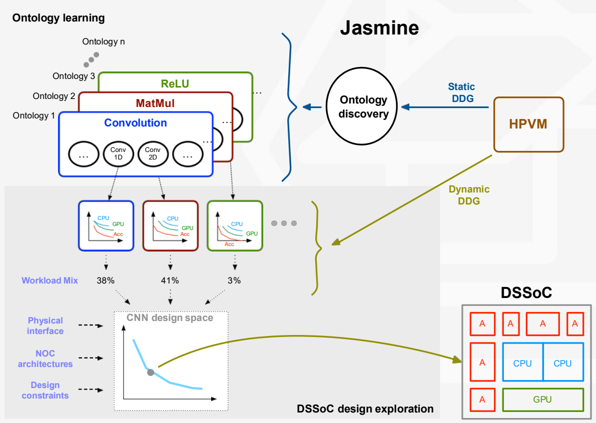


Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

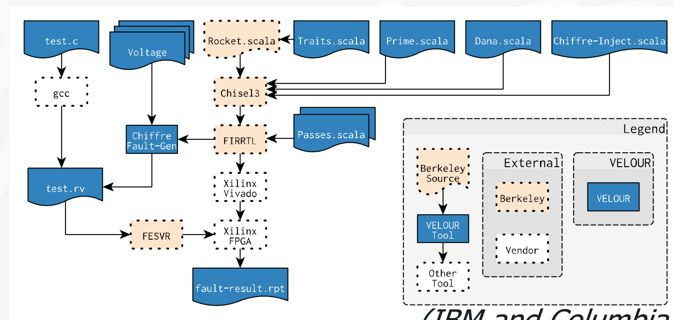
EPOCHS: KEY FOCUS AREAS

- Systematic identification of accelerator components

(Harvard University)



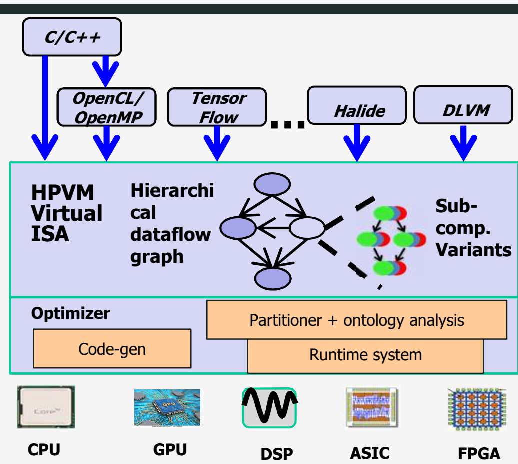
- Agile open-source hardware design flow from application to silicon



(IBM and Columbia)

- Open-source compiler and programming model

(UIUC)



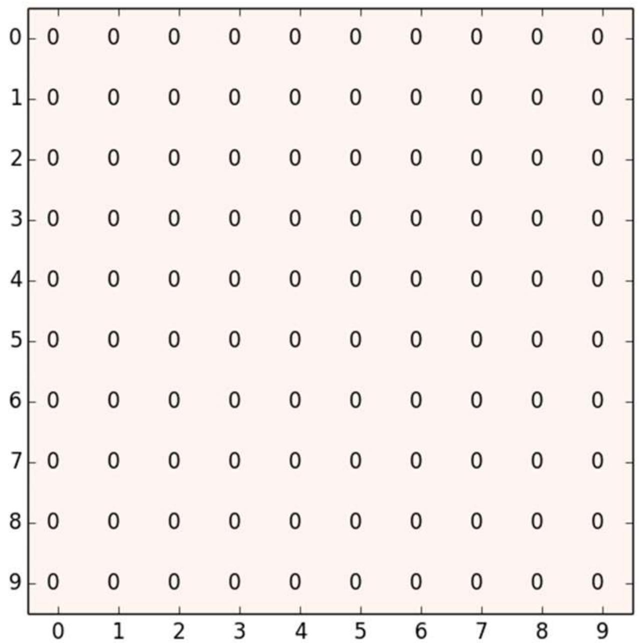
- Distributed "swarm" resource management
  - Environment:
    - Core-level voltage domains (synchronous)
    - Very low voltage (VLV) domains
    - Asynchrony between voltage domains
    - Per-core power gating
    - Header/footer array on 3D or interposer
  - Decentralized self-management
    - Manage performance, power / temperature, resilience
    - Decentralized error recovery, reconfiguration
    - Continuous, adaptive learning with core-to-core knowledge transfer/update

(IBM and others)

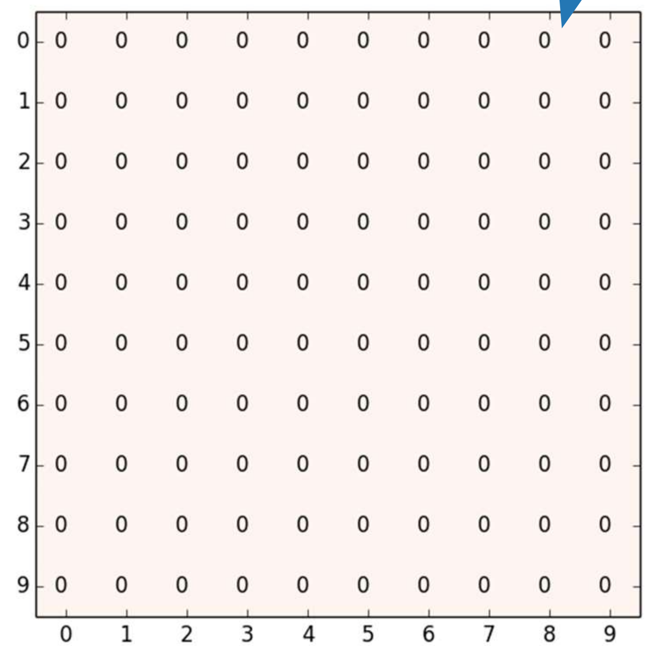
TOKEN BASED SWARM POWER MANAGEMENT SIMULATION

Numbers in each core represent the deficit of tokens
(the lower the better)

Disabled*



Enabled*



** Animation frames taken every 100 simulation iterations*

SUMMARY

- **The need for SoC specialization is increasing, but costs are prohibitive and increasing**
  - A new approach to realizing novel, heterogenous architectures is needed

- **EPOCHS goal:** improve cognitive IoT efficiency at the edge by 100X within 5 – 7 years
- **EPOCHS "super" domain:** embedded systems for intelligent, networked autonomous vehicles
  - Domains: computer vision, software defined radio

- **Key innovation areas:**
  - Decentralized hardware control assist to support OS task scheduler
  - Re-targetable, adaptive compiler
  - Formalism for workload analysis
  - GALS MAC layer with autonomous management
  - Design ecosystem supporting wide-operating-range, heterogeneous SoCs, scalable from 25W down to 5W

THANK YOU!

Photo by Balthazar Korab
Source: http://www.shorpy.com/node/15488

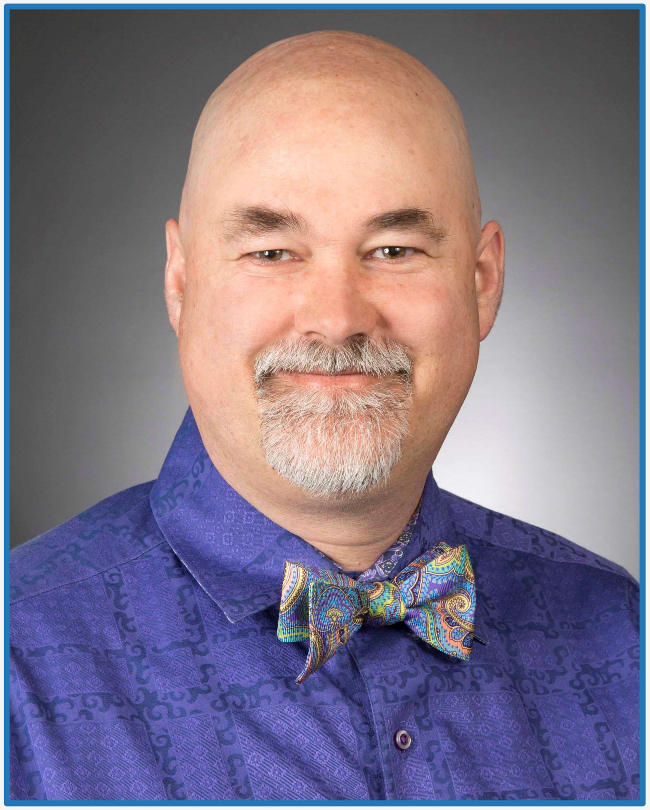
Jeff Burns
jlburns@us.ibm.com

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

ERI ELECTRONICS RESURGENCE INITIATIVE

SUMMIT

2018 | SAN FRANCISCO, CA | **JULY 23-25**



DANIEL BLISS

ARIZONA STATE UNIVERSITY
PROF. & DIRECTOR OF WISCA

FROM DOMAIN TO CHIP TO IMPLEMENTATION

PROF. DANIEL W. BLISS

ELECTRICAL, COMPUTER, AND ENERGY ENGINEERING

ARIZONA STATE UNIVERSITY

DIRECTOR OF WISCA

CENTER FOR WIRELESS INFORMATION SYSTEMS AND COMPUTATIONAL ARCHITECTURES

WISCA Center for Wireless Information Systems and Computational Architectures

ASU Arizona State University

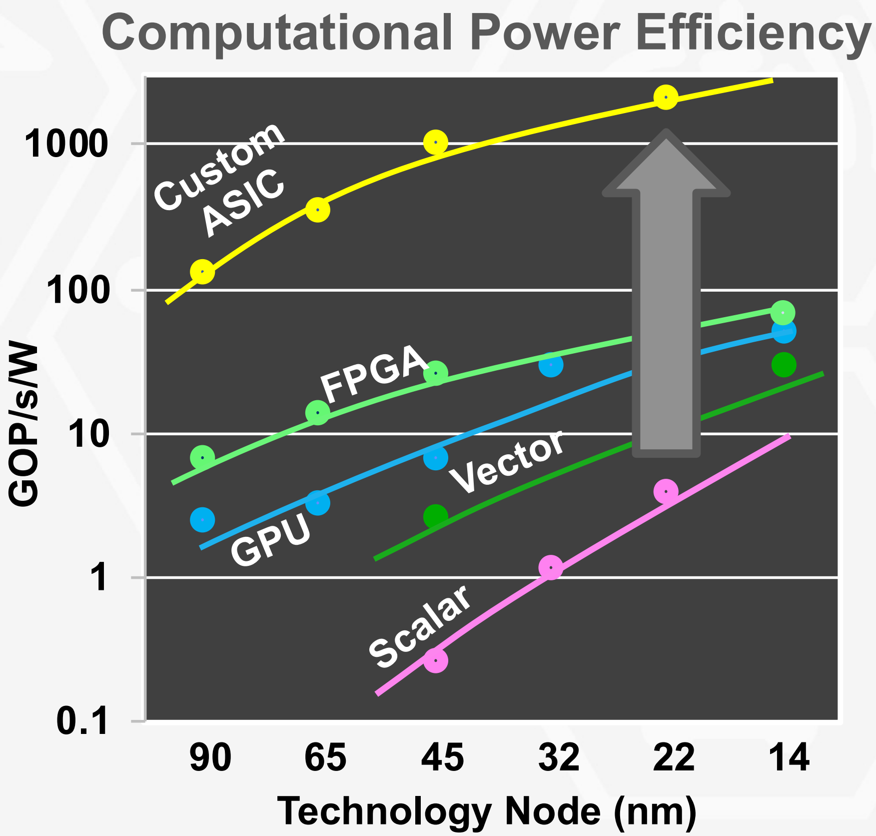
TOPICS

- What do we need?
- How do we do it?
- Who cares?

WHAT DO WE NEED?

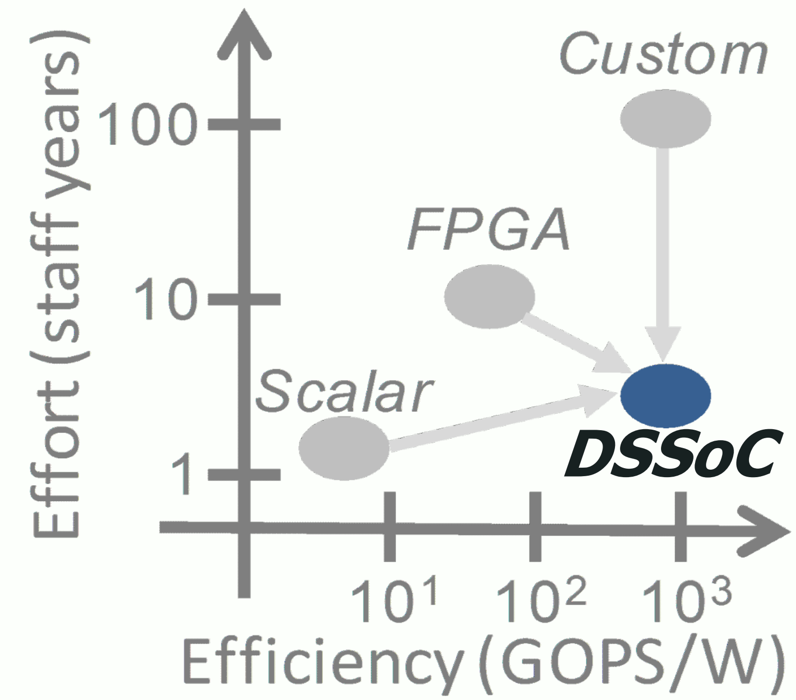
SIGNAL PROCESSING TECHNOLOGY TRADE

- Imposes trade between flexibility and performance
- Limits high-performance systems to rigid and well funded applications
- Affects system designers view of what is possible



DSSOC GOALS

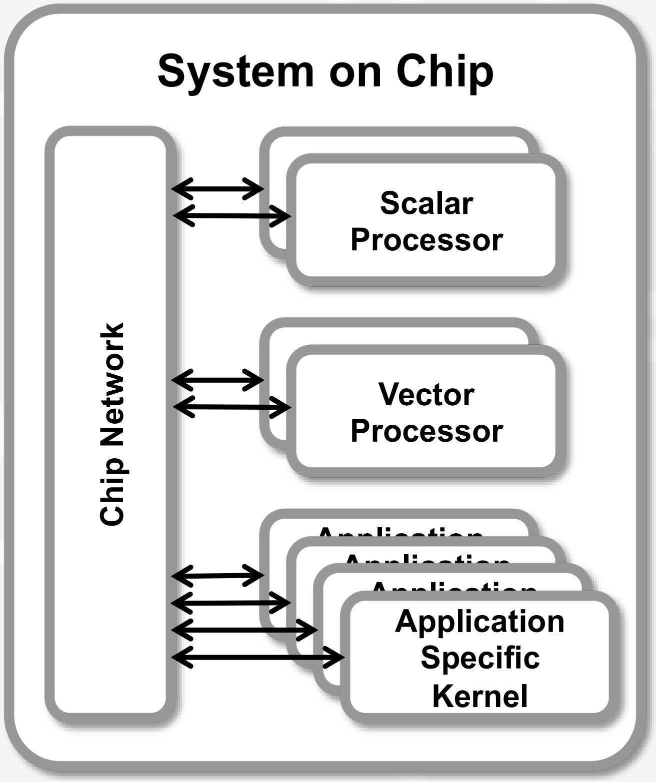
- Break traditional difficulty-performance trade
- Enable efficient design and use of processors
- Provide tools to enable efficient domain-specific chip design
- Provide tools to enable easy implementation



HOW DO WE DO IT?

ADDRESS HETEROGENOUS PROCESSOR CHALLENGES

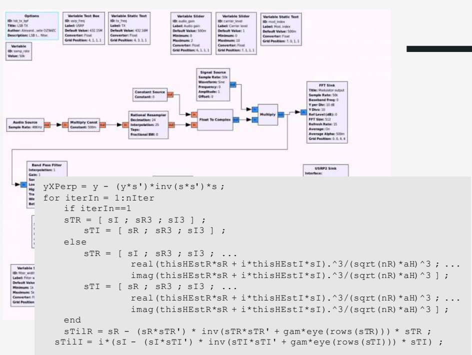
- Use heterogenous processors
- Introduced new set of problems to address
- Specify application-specific kernels
- Aid chip design and integration
- Reduce programming effort
- Optimize run-time processor and network utilization for multiple simultaneous applications

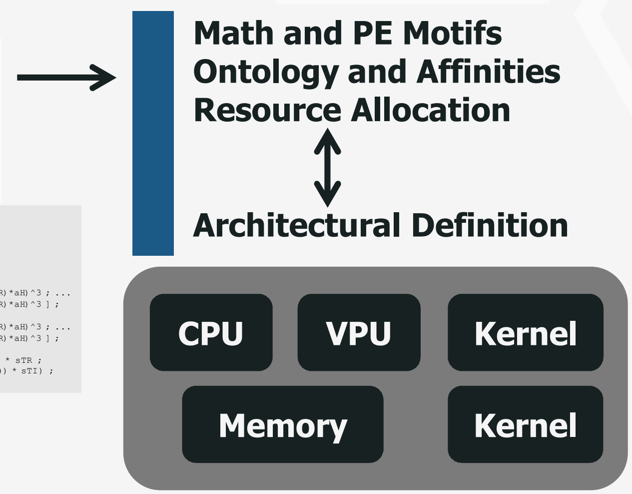


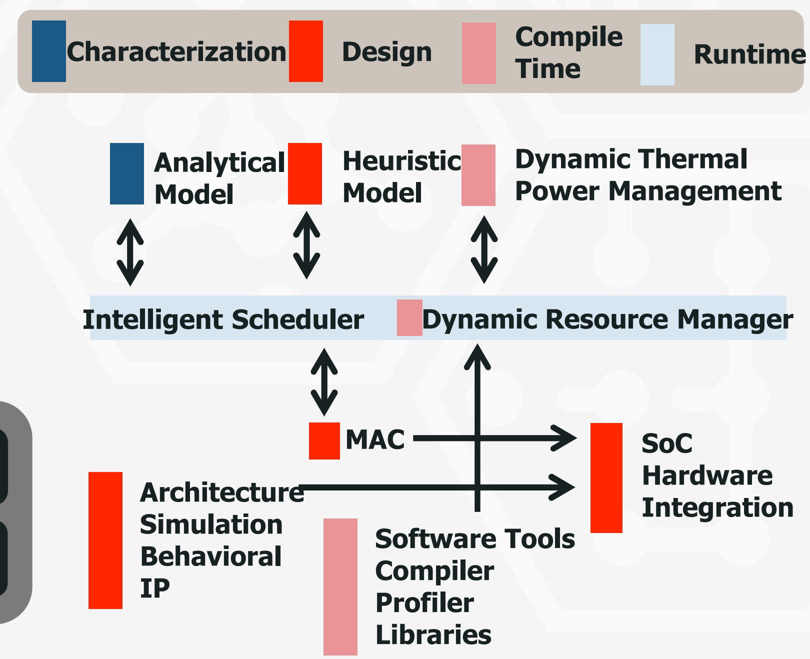
DESIGN AND INTEGRATION

- Integrate across time scales
- Integrate learning at multiple levels

Domain-Focused Advanced Software-Reconfigurable Heterogeneous (DASH) - SoC







WISCA Center for Wireless Information Systems and Computational Architectures

DASH TEAM

- Built team of experts to address all system aspects
- Bring significant academic, commercial, and DoD experience

Prime
ASU

Ontology
ASU

Software
University of Michigan
ARM
University of Arizona

IS
ASU
Carnegie Mellon
EpiSci
University of Arizona

MAC
ASU
Carnegie Mellon

Hardware
University of Michigan
ARM

Transition
General Dynamics

WHO CARES?

SYSTEM PLATFORM LIMITATIONS

- Address needs of proliferation of smaller platforms
- Limits power to ~ 100mW to 10 W
- Excludes use of 19” rack of servers

New Military Radio Functions

Military Robots

Multifunction UASs

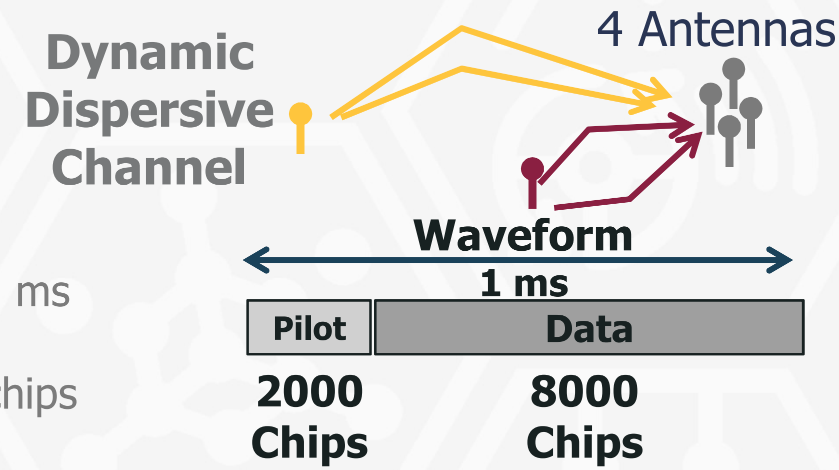
Advanced Medical Devices

ADAPTIVE INTERFERENCE-MITIGATION COMMUNICATIONS

- Mitigate interference with array in dynamic dispersive environments
- Design communications receiver for
  - Bandwidth = 10 MHz → t_{chip} = 0.1 ms
  - Coherence Interval ~ few ms → $t_{waveform}$ = 1 ms
  - Delay Spread ~ 500 ns → n_{taps} ~ 20
  - INR = 30 dB, SNR = 10 dB → n_{pilot} > 1000 chips
  - Oversampling = 2
- Implement space-time beamformer

$$\mathbf{w} = \hat{\mathbf{R}}^{-1}\,\mathbf{v} = (\tilde{\mathbf{Z}}\,\tilde{\mathbf{Z}}^{\dagger})^{-1}\,\tilde{\mathbf{Z}}\,\underline{\mathbf{u}}^{\dagger}_{\text{pilot}} \in \mathbb{C}^{(n_{\text{ant}}\cdot n_{\text{tap}}\cdot n_{\text{oversamp}})\times 1}$$

  - Dominant cost estimating is $\hat{\mathbf{R}} \in \mathbb{C}^{(n_{\text{ant}}\cdot n_{\text{tap}}\cdot n_{\text{oversamp}})\times(n_{\text{ant}}\cdot n_{\text{tap}}\cdot n_{\text{oversamp}})} \rightarrow \underbrace{\mathbb{C}^{160\times 160}}_{\sim 25{,}000 \text{ Complex Values}}$

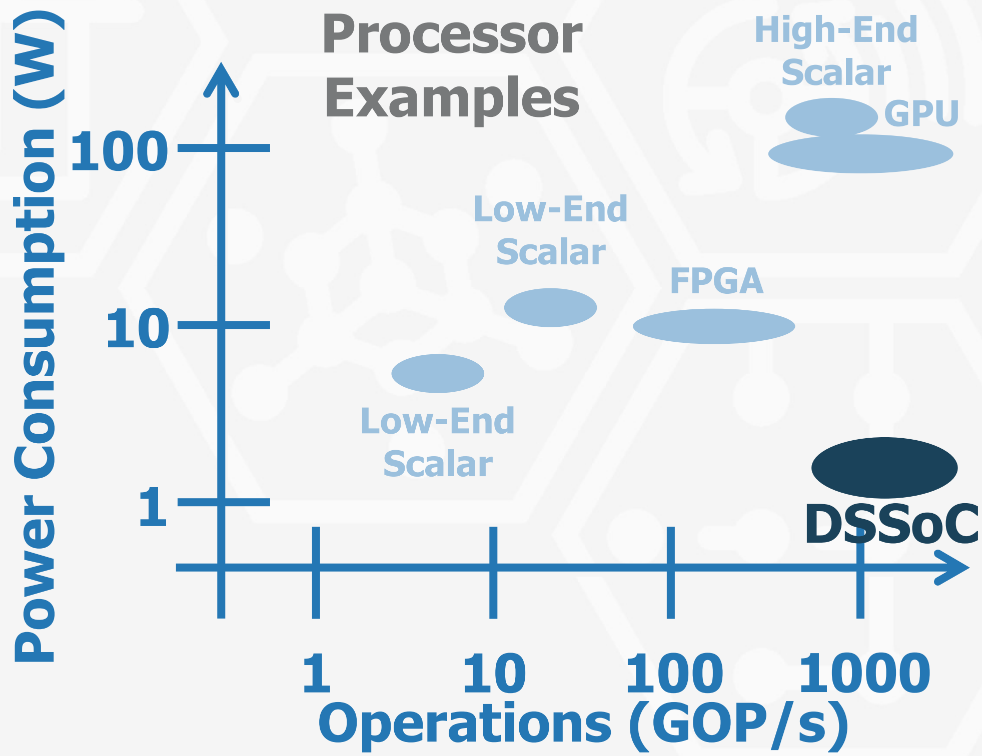


$$\mathcal{O} \sim k\,(n_{\text{ant}}\cdot n_{\text{tap}}\cdot n_{\text{oversamp}})^2\, n_{\text{pilot}}\, n_{\text{oversamp}}/t_{\text{waveform}}$$
$$\approx 1\ \text{TOP/s}$$

WISCA Center for Wireless Information Systems and Computational Architectures

EXAMPLES OF CURRENT PROCESSORS

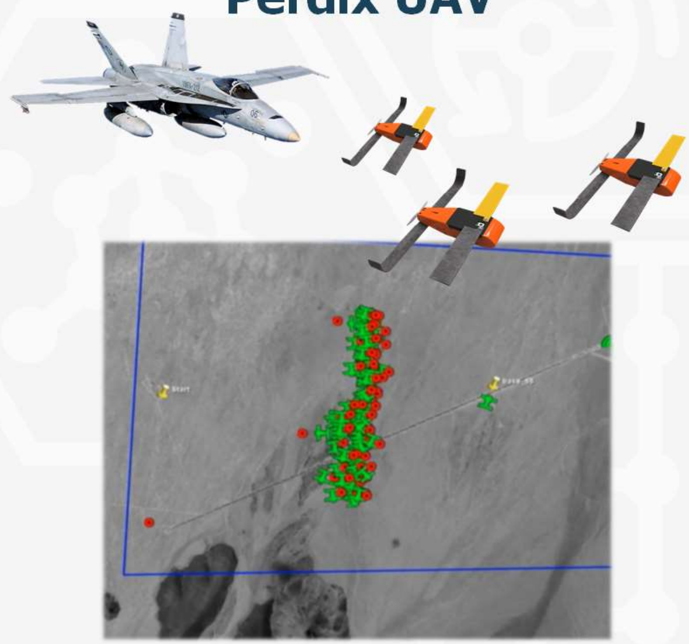
- Place current technology relative evolving system requirements
- Consider signal processing applications
- Clarifies need for DSSoC approach



UAS SWARMS

- Consider pod-deployable UAV systems
- Employ for distributed SIGINT, EW, radar, or communications
- Address limited avionics power
  10s of Watts
- Translate from amusing to viable

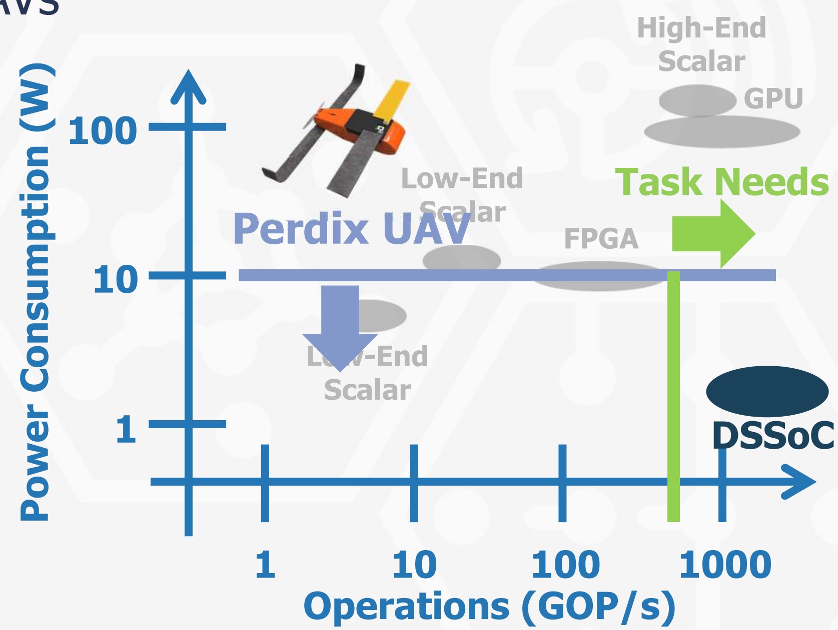
Pod-Deployable Perdix UAV



WISCA
Center for Wireless Information Systems and Computational Architectures

ADVANCED UAV APPLICATIONS

- Enable advanced applications on small UAVs
- Address computation needs
  - Attractor of 1000 GOP/s tasks
  - Advanced SIGINT
    Modern communications ~ 1000 GOP/s
  - Bistatic SAR
    Match filter ~ 1000 GOP/s, TB = 1000
  - Advanced video processing
    8k video image alignment ~ 1000 GOP/s



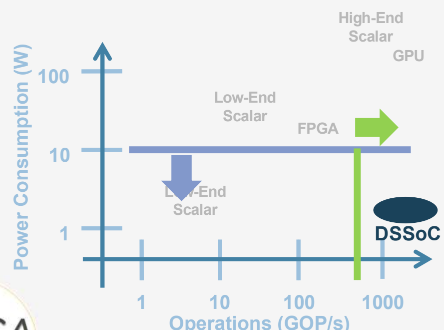
WISCA
Center for Wireless Information Systems and Computational Architectures

ADVANCED UAS POSITIONING

- Perform precision positioning and communications while using resources
- Employ MIMO phase-accurate ranging limited spectral
- Requires significant multithreaded computational capabilities

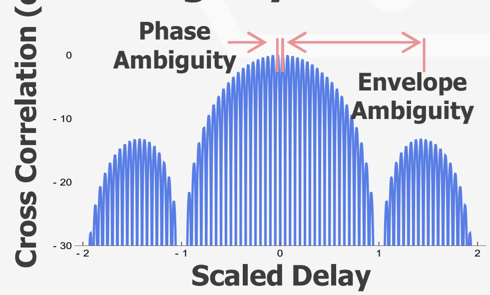
~ 100 GOP/s per node x *10s of nodes*

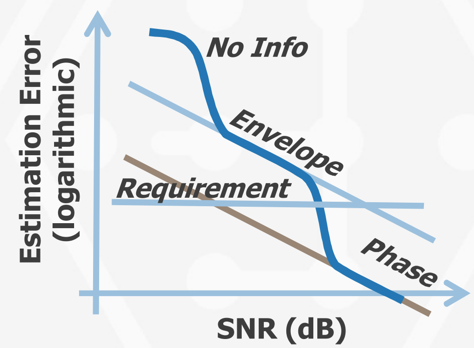
UAS Remote Positioning



Theoretical Ranging Performance

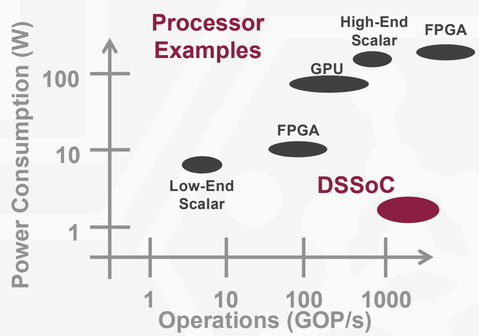
Ambiguity Function



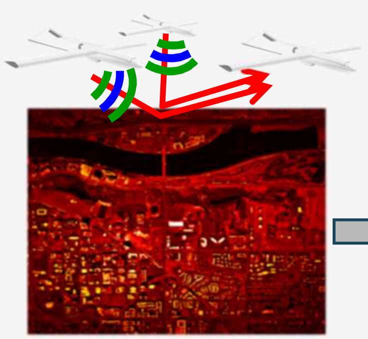


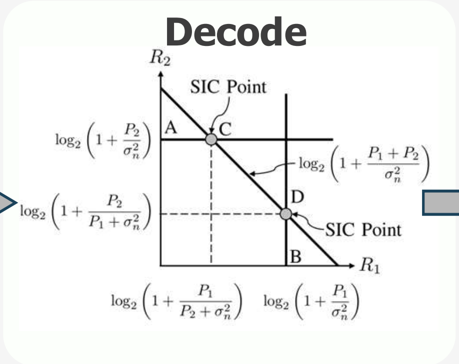
RF CONVERGENCE COMMUNICATIONS & SAR

- Perform joint communications and multistatic SAR by using multifunction waveforms
- Requires significant multithreaded computational capability *~ 10s TOP/s*



Multi Access SAR





Re-Modulate

Synthetic Aperture Processing

SUMMARY

- Motivated need to address evolving requirements
- Defined new SoC development and implementation goals
- Provided examples of advanced applications that will benefit from DSSoC capabilities